An algorithm to relate surface roughness with local geometry of protein exterior.

Anirban Banerji

Bioinformatics Centre, University of Pune,

Pune-411007,Maharashtra, India (anirbanab@gmail.com)

Indira Ghosh*

School of Information Technology, Jawaharlal Nehru University

New Delhi-110067, India

indirag@mail.jnu.ac.in

Date : 20^{th} December, 2008

Abstract

Changes in the extent of local concavity and surface roughness of binding sites of proteins have long been considered as useful markers to study functional sites of proteins. However, an objective algorithm that describes the connection between the simultaneous changes of these important parameters, eludes the students of structural biology. Here we propose a simple yet rigorous algorithm that attempts to achieve the same. A generalized index is constructed here that not only describes the process but also reduces the estimation error in local shape characterization. The proposed methodology is easier to implement with computational tool-set.

Introduction :

The concavity of the active sites of enzymes and the changes in their local geometry has long been recognized as key features to study their structure-function relationship (Lewis 1991; Pettit et al. 2007). Alongside this recognition, came the acceptance that change in surface roughness of patch of protein surface can serve as an equally useful marker, at least when attempting to find functional sites of that protein (Lewis and Rees 1985; Pettit and Bowie 1999). While there are works that successfully quantify important aspects of geometry of local shape of protein exterior (Laskowski et al. 1996; Peters et al. 1996; Liang et al. 1998; Brady and Stouten 2000); an algorithm that relates the local shape change to change of roughness of surface patch that holds this shape, is extremely difficult to find. The reason for this conspicuous absence can probably be attributed to the complex nature of functional dependencies that influence surface roughness of any given patch of protein surface. It is owing to these (seemingly innumerable) dependencies that the surface roughness appears to be an $n-$dimensional random variable. In this paper, we propose a simple yet rigorous algorithm that transforms the surface roughness description from that of a $n-$dimensional random variable to a system of simple algebraic equations. Further, we define a generalized index that captures the simultaneous changes in local geometry and the roughness of the patch of protein surface that engulfs it. It can immediately be recognized that the algebraic description of this complicated dependency problem is much easily implementable in computational form than the same in the realm of $n-$dimensional continuous random variables.

Methodology :

Since the entire phenomenon of enzyme-inhibitor interaction (like most of the biological interactions) is time-dependent and context-dependent, the local shape

of active site of enzyme is expected to undergo certain small yet significant change in its topography. Further, since any shape is necessarily coated with a surface, a minute change in the shape can account for a subtle change in the magnitude of surface roughness (at least potentially) too. Hence to describe the situation objectively, let $[s = 0, 1, 2, \ldots, r]$ be an estimate of the local shape of the active site of an enzyme when it is undergoing interaction with some inhibitor protein; and let $[F = (f_1, f_2, \ldots, f_n)]$ be the corresponding vector of the state representing surface roughness (expressed with fractal dimensions) of the patch of the surface containing the aforementioned local shape. We define a function $\phi(F)$, on which the functional $Z = \|\phi(F) - s\|$ attains a minimum. Since the magnitude of the functional $Z = \|\phi(F) - s\|$ has the capability to describe a change in either $\phi(F)$ or s; the minimum magnitude of the functional will imply a state where the extent of change in $\phi(F)$ follows the trend in extent of change in s, in the closest terms. Hence it is this minimum magnitude of the functional Z that we define as generalized index for studying the change in local shape of the enzyme active site, because only with this magnitude of functional Z, the change in both shape and surface roughness can be simultaneously described in the best objective manner.

It is clear that the solution of the give problem is determined by the choice of the norm $\|.\|$. To describe the situation from an unbiased perspective, we have considered F as an $n-$dimensional continuous random variable with correlation matrix C. We could then define $\phi(F)$ as a predictor of a random variable s having a minimum mean square value. Thus the functional Z was re-written as :

$$Z = \sum_{s=0}^{r} p_s E\left[\left(\phi\left(F\right) - s\right)^2 |s\right] \tag{1}$$

where E represented the expectation operator and p_s denoted the probability of occurrence of any arbitrarily chosen state of the shape, s. In the case where $E\left(\phi\left(F\right)|s\right) = s$ for all $[s = 0, 1, \ldots, r]$, the given functional state attains a minimum.

To describe the situation from a bottom-up perspective, i.e., describing with respect to an individual fractal dimension (f_k) that represented a particular value of surface roughness; we had : $E\left(f_k|s\right) = s$ for all s and $[k = 0, 1, \ldots, r]$. We assume here that the form of the relations $E(f_i|s)$ and the coefficients α_i, $(i = 1, 2, ..., n)$ remain unchanged in time.

It was then possible to represent $\phi\left(F\right)$ as a linear combination of indices :

$$\phi\left(F\right) = \sum_{i=1}^{n} \alpha_i f_i \tag{2}$$

where $\sum_{i=1}^{n} \alpha_i = 1$; since in this case $E\left(\phi\left(F\right)|s\right) = s$. The coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ were evaluated from the minimum condition of the Lagrangian :

$$\widetilde{Z}\left(\alpha, \lambda\right) = \sum_{s} p_s E\left[\left(\sum_{i=1}^{n} \alpha_i \left(f_i - s\right)\right)^2 |s\right] + \lambda\left(\sum_{i=1}^{n} \alpha_i - 1\right) \tag{3}$$

Differentiating $\widetilde{Z}\left(\alpha, \lambda\right)$ with respect to α_i $(i = 1, 2, ..., n)$ and equating the derivatives to zero we obtain a system of algebraic equations in α and λ :

$$\sum_{i=1}^{n} \alpha_i \sum_{s} p_s E\left[\left(f_k - s\right)\left(f_i - s\right)|s\right] + \lambda = 0 \tag{4}$$

where $(k = 1, 2, ..., n)$ and $\sum \alpha_i = 1$.

In case where C is the diagonal matrix, the solution of the given system was written as :

$$\alpha_i = \left(\sum_{k=1}^{n} \prod_{j \neq k} C_{jj}\right)^{-1} \prod_{k \neq i} C_{kk} \tag{5}$$

and

$$\lambda = -\left(\sum_{k=1}^{n} \prod_{j \neq k} C_{jj}\right)^{-1} \prod_{k=i} C_{kk} \tag{6}$$

further, for the variance of the generalized index we get

$$E\left(\phi\left(F\right) - s\right)^2 = \left(\sum_{k=1}^{n} \frac{1}{C_{kk}}\right)^{-1} \tag{7}$$

Since C_{kk}, $(k = 1, 2, ..., n)$, are positive and bounded, we arrived at :

$$\lim_{n \to \infty} E\left(\phi\left(F\right) - s\right)^2 = 0. \tag{8}$$

Eq^n-8 shows unambiguously how the introduction of the generalized index could reduce the estimation error in local shape characterization and how such an index made it possible to transform the problem from the complex analysis of an

$n-$dimensional variable to that of involving scaler parameters. This particular aspect of the present work makes it easily compatible with the computational tool-kit. It is achieved without compromising with the necessary rigor, essential to describe the intricate (time-dependent and context-dependent) coupling between simultaneous changes of two variables with large dependencies.

In a special case, the regressions $E\left(f_k|s\right) = \psi_k\left(s\right)$, (k = 1, 2, ..., n) can be non-linear. This can be reduced to the one considered by transforming regressive relations to a linear form; for example, with the help of the polynomials $\mu_k\left(f_k\right)$, $(k = 1, 2, ..., n)$; so that $E\left[\mu_k\left(f_k\right)|s\right] = s$. Then the generalized index will be linear combination of polynomials :

$\phi(F) = \sum_{i=1}^{n} \alpha_i \mu_i(f_i)$.

Conclusion :

However, simple and useful as this technique is, it is not immune to limitations. For example, the algorithm proposed in the current work might not be efficient when applied to cases where functional sites are not appreciably concave. Progressing on similar line thinking, since the protein-protein interaction interface is known to be planer (Jones and Thornton 1997), the scope of the application of an algorithm like the one proposed, might be restricted. Having said that, since for many proteins, the functional sites are known to be containing concave shapes (in the form of clefts and pockets) (Laskowski et al. 1996), the spectrum of contexts where the present algorithm can be applied assumes an impressive range. The form of simple algebraic equations, with which the final description of this complex biophysical interaction is expressed, makes it easily implementable with elementary computational apparatus.

Acknowledgment : This study was supported by COE scheme, Department of Biotechnology, Government of India.

Reference :

*) Brady GP Jr, Stouten PF (2000) Fast prediction and visualization of protein binding pockets with PASS. J Comp-Aided Mol Des. 14 : 383–401.

*) Jones S, Thornton JM. (1997) Prediction of protein-protein interaction sites using patch analysis. J Mol Biol 272 : 133–143.

*) Laskowski RA, Luscombe NM, Swindells MB, Thornton JM (1996) Protein clefts in molecular recognition and function. Protein Sci. 5: 2438–2452.

*) Lewis M, Rees DC (1985) Fractal surfaces of proteins. Science 230: 1163-1165.

Lewis RA, (1991) Clefts and binding sites in protein receptors. Methods Enzymol. 202:126–156.

*) Liang J, Edelsbrunner H, Woodward C. (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci. 7:1884–1897.

*) Peters KP, Fauck J, Frommel C. (1996) The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. J Mol Biol. 256: 201–213.

*) Pettit FK, Bare E, Tsai A, Bowie JU (2007) HotPatch: A Statistical Approach to Finding Biologically Relevant Features on Protein Surfaces. J Mol. Biol. 369: 863–879.

*) Pettit FK, Bowie JU (1999) Protein surface roughness and small molecular binding sites. J Mol. Biol. 285: 1377–1382.